Horse kicks, flying bombs and potsherds: statistical theory contributes to archaeological survey

Clive Orton

In the application of statistics in archaeology, it is essential to choose an appropriate statistical model for the specific archaeological problem being addressed. This article discusses the application of statistical models to large-scale archaeological field survey and describes the development of software (sherdnav) designed to guide fieldwalkers to where they should look next, once they have surveyed all the primary spots in a block, and its application to the Noviodunum Archaeological Project (NAP).

During a career spent in developing, using and teaching the applications of statistical techniques in archaeology, it has been apparent to me that a crucial first step towards a successful analysis is the choice of an appropriate statistical model to

and how it led to the development of software that could be used in "real time" in the field

• Present some preliminary results and show how they can be interpreted

• Sketch out a possible route for future developments.

Figure 1 *Location map showing site of the Roman fort of Noviodunum.*

describe the situation being studied. The process of choosing a model not only focuses the mind on the essentials of a problem, it also guides one towards an appropriate choice of technique, which in these days of user-friendly software can be the most difficult part of an analysis. This has been a particularly valuable exercise in two case studies: an analysis of Museum Collection Condition Surveys[1] and a study of ceramic production centres.[2] More recently, the development of a range of statistical models of increasing complexity has proved its value in a new area – the undertaking of large-scale archaeological field survey, and in particular the Noviodunum Archaeological Project (NAP) (Fig. 1).[3]

The aims of this paper are to:

• Outline the problem of carrying out archaeological field surveys effectively and efficiently

• Tell the story of the search for models that could guide the progress of the fieldwork,

The problem[4]

One aim of the NAP is to locate and characterize archaeological sites in the area surrounding Noviodunum, to shed light on the ways in which the fort may have interacted with its hinterland. This is achieved by extensive field-walking, that is by walking over the terrain and collecting surface scatters of artefacts which may indicate the locations of buried sites. Traditionally, this is done by teams of walkers moving linearly across the landscape at a pre-determined spacing (e.g. 10m apart).[5] However, the density of surface artefacts around Noviodunum is so great that this would have rapidly filled the available storage space, so a less intensive approach, the "spot" method, was used. In this approach, a regular grid of circular "spots" of a chosen size is laid out, and all the artefacts within each spot are collected; nothing is collected from outside the spots. At Noviodunum, the spots were given an area of 2 sqare metres (i.e. a radius of 0.80m) and were placed on a rectangular grid at 30m intervals. This is achieved by laying out a central line of points using a total station, and then laying out the points either side of that line using an optical right-angle and tapes. The team of five walkers progresses along the transect, stopping every fifth line to input the data to the program, and undertake any extra collection that is indicated (see below). The transects are located in the landscape using hand-held GPS (Figs 2–4).

It was anticipated that some spots would contain a high density of artefacts (specifically, Roman pottery sherds), and would be considered "on site", while others would contain a lower density

Figure 2 *Spot walking at Noviodunum. From top-centre clockwise: laying out the centre line using a total station; setting out the line of spots using an optical right angle; locating the end of a transect using a handheld GPS; a line of spots being walked.*

of sherds and be considered "off site".[6] It would then be possible to define roughly the extent of any sites that had been located by simply drawing a curve round contiguous groups of high-density spots. Because of the wide (30m) spacing between the spots, this would give only a rough delineation of any sites, and a finer drawing of their outline was felt to be necessary. It was decided to achieve this by surveying further spots, at 10m intervals, but only when needed to refine the boundary of a site. Considering pairs of spots 30m apart, there are three possible situations and two practical outcomes, as follows:

a) both spots are "high density": both are "on site", so there is no need to survey between them

b) both spots are "low density": both are "off site", so there is no need to survey between them

c) one spot is "high density" and the other is "low density": one is "on site" and one is "off site", so there is a need to survey two further spots between them, reducing the spacing from 30m to 10m.

This approach throws down the statistical challenge of deciding which spots are of "high density" and which are of "low density", preferably with the minimum of data processing so that any intermediate spots could be surveyed as soon as possible. This was technically possible because each team has a PDA and could thus enter data, carry out calculations and ascertain which (if any) further spots needed to be surveyed.

The search for a model

For purposes of analysis, the spots are grouped into "blocks" of 25 spots (five-by-five). From a statistical point of view, the problem is to take a set of 25 counts (the numbers of sherds in each of the 25 spots in a block), and to decide whether they could all reasonably be seen as samples from the same background density (accepting that some variation will occur), or whether this simple belief is untenable, and they must therefore be from zones of different densities. The first alternative corresponds to options (a) and (b) above, and the second to option (c).[7]

This begs a further question – if there were sherds scattered randomly across a block, whether at high or low density, what sort of pattern of variation would we expect to find between the numbers found in each spot? Let's look at the sorts of numbers that might be involved, drawing on data from the 2005 survey. Fig. 4 shows the numbers of Roman sherds found in each spot of a test block. It's useful to re-cast this as a histogram, which shows how many spots contain no sherds, one sherd, two sherds, etc. (Fig. 5). How might this histogram look if sherds had been scattered randomly in the block at the same overall density? Does our histogram look anything like this theoretical ideal (which is known as complete spatial randomness or CSR)?

At this point we need to digress into statistical theory. In 1837 a French mathematician, Siméon-Denis Poisson, described a mathematical model for the numbers of "events" that would occur in a certain "interval", if there were very many such events, each of which had a very small probability of occurring in any particular interval.[8] This became known, not surprisingly, as the Poisson distribution, and was admired as a piece of mathematics, but remained an abstract concept until 1898, when Ladislaus von Bortkiewicz, a German economist, interpreted "event" as a soldier in the Prussian army being killed by a horse-kick, and the "interval" as a period of time, such as a year. His data on the numbers of Prussian soldiers killed in this way did indeed fit the theoretical Poisson distribution, i.e. deaths occurred randomly over time.[9] Later, it was appreciated that "interval" could refer to space as well as time. The British actuary Clarke was able to fit a Poisson model to the spatial distribution of flying-bomb hits on London, demonstrating that within the general target area their spatial occurrence was random.[10] In a further
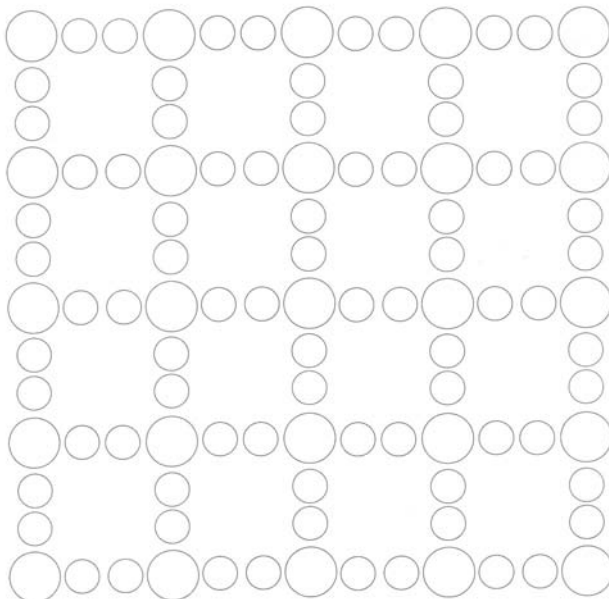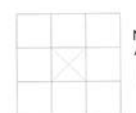

Figure 3 *Recording sheet for one survey block of the NAP.*
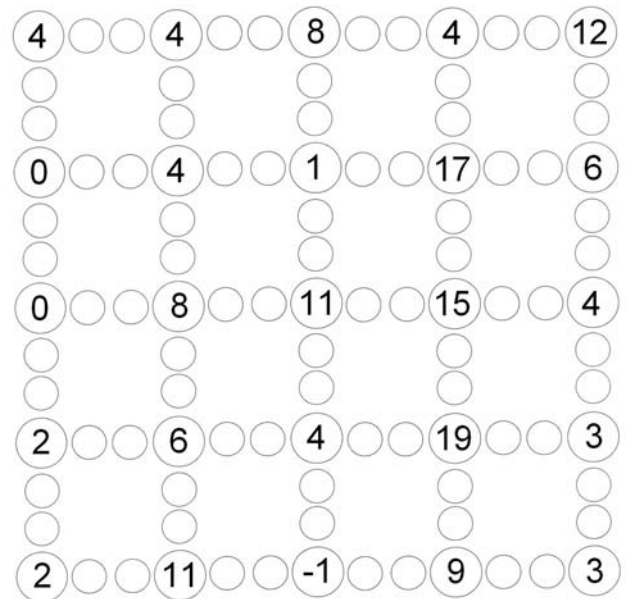
Figure 4 *Completed recording sheet for the test block, showing numbers of sherds found in each spot (a count of -1 indicates a spot that could not be surveyed).*
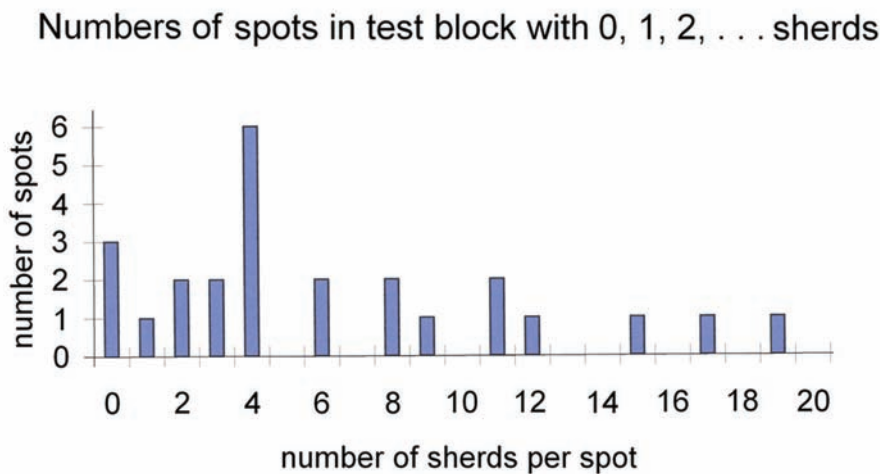
Numbers of spots in test block with 0, 1, 2, . . . sherds

Figure 5 *Histogram of the data from Figure 4, showing the numbers of spots with 0, 1, 2, …
sherds.*

example, O'Sullivan and Unwin showed that the distribution of coffee shops of a particular chain in central London in 2000 did not fit a Poisson model, and so was not spatially random.[11]

Software for fieldwalkers: *sherdnav*
What works for horse-kicks, flying bombs and coffee shops, should work for Roman potsherds. From this idea was born *sherdnav*: software to guide fieldwalkers to where they should look next, once they have surveyed all the available spots in a block. *Sherdnav* starts by testing whether the counts of sherds can be fitted to a Poisson (P) distribution; if they can, their block is in situation (a) or (b). If that does not fit, a more complex model, PP (double Poisson, consisting of a mixture of two Poisson distributions with different means, in unknown proportions) is fitted, corresponding to situation (c). The means of the two distributions, and their relative weights, are estimated using maximum likelihood estimation (mle).[12] If the lower of the two means is zero, we have the P+ model (Poisson plus a number of zeros)[13] as a special case of the PP model. So far, the PP model has never failed to fit.[14] *Sherdnav* can find both the "simplest adequate model" and the "best-fit model"[15] to describe the counts in a block.

If the PP model is needed, the block is partly on-site and partly off-site. *Sherdnav* then compares the count in each spot with those in its immediate neighbours; if they belong to different distributions, then survey of the intermediate (10m) spots is recommended (intensive survey). If the block appears to be partly or wholly on-site, survey of the appropriate neighbouring squares is recommended (extensive survey).

Figure 6 shows the division of a test block into spots that are on-site and spots that are off-site, and indicates which intermediate spots need to be surveyed

intensively. Blocks adjacent to sides of the block which are wholly or partly on-site are recommended for extensive survey.

Implementation

Sherdnav is designed to run on Pocket Excel© on a PDA. Users enter data in three adjacent columns: *eastings*, *northings* and *count*, using -1 to indicate unsurveyed spots. Spots are entered in a standard order. The outcomes consist of four additional columns headed north, east, south and west, telling users whether they should survey the extra spots immediately to the north, east, south and west of each spot, and a row telling them whether they should survey the blocks to the north, east, south and west. All the calculations and intermediate working are hidden from users, thus making good use of the small screen.

Use in the field
The software was first used in the field in April 2006, revealing weaknesses that led to some false positives (surveying spots that did not need it). These were put right for the summer 2006 season, which unfortunately discovered very little,
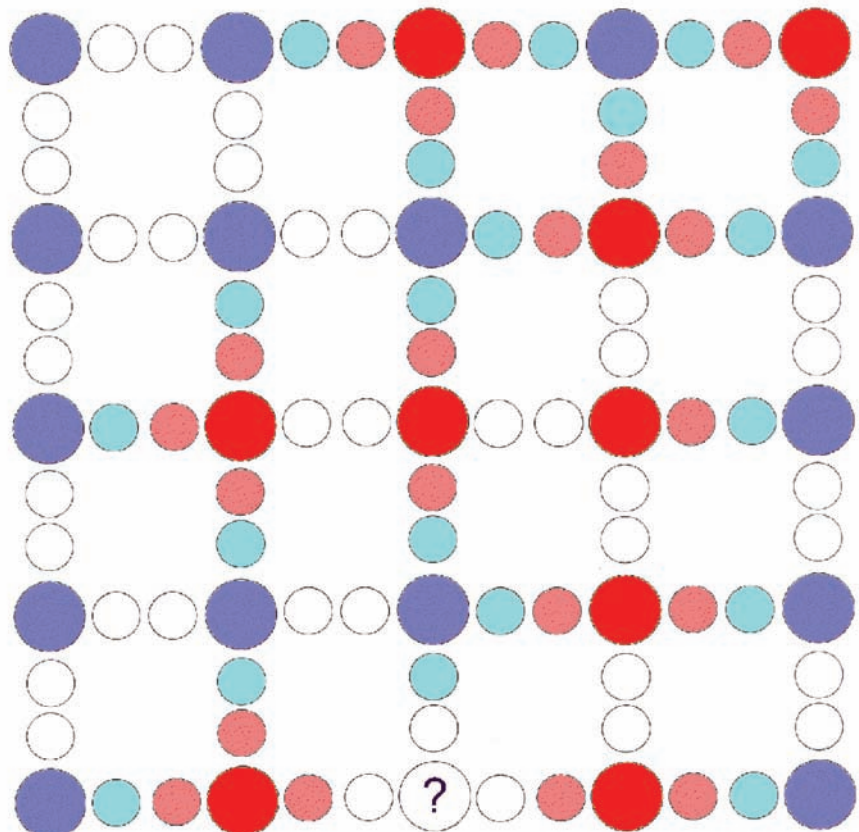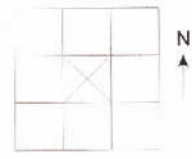

Figure 6 *Recording sheet for test/δ block after analysis, showing on-site spots (red), off-site spots (blue)
and extra spots to be surveyed (shaded).*

and did not seriously test the software's capabilities. The software worked very well during the Easter 2007 survey, and the data are currently being evaluated.

Future work

The main criticism of the *sherdnav* approach is that it is, as currently implemented, purely local. It gives clear advice about actions within a block, but says nothing about what to do in the 30m-wide strips between blocks. Indeed, the definitions of high and low density could be quite different in two adjacent blocks. Additional software is needed to sit above the level of the individual blocks, to allow them to be metaphorically stitched together. This is the next major task. Other lesser tasks are to examine the relative merits of the "simplest adequate model" and the "best fit model", and to examine whether the theoretical advantages of mle over MoM outweigh the practical complications.

Notes

1. In this case, the overall condition of a museum collection was represented by a probabilistic model known as a Markov Chain, see C. Orton, "Markov models for museums", in *Interfacing the Past. Computer Applications and Quantitative Methods in Archaeology*, H. Kamermans and K. Fenema (eds),149–53, CAA95 Analecta Praehistorica Leidensia 28,1996).

2. The use of a multivariate normal model to describe the operation of a ceramic production site was described informally in C. Orton, "The sherds tell a story: Roman rural potters in the London area", *Archaeology International* (2002/2003) 18–20; the mathematical details were given in C. Orton, "Never under-estimate the power of a model", in *Archaeological Informatics: Pushing the Envelope,* G. Burenhalt (ed.), 495–9, CAA2001 Computer Applications and Quantitative Methods in Archaeology (Oxford: BAR Int Ser 1016, 2002). This approach was later used in a more general sense to help discuss the scale of pottery production in the medieval Russian city of Novgorod, see C. Orton, "Indicators of craft specialisation in medieval ceramics from north-west Russia", in *The Archaeology of Medieval Novgorod and its Territory: A study of centre/periphery relationships based on the archaeological evidence for the exploitation and processing of natural resources within Novgorod, its hinterland and the northeast fringes of its territory,* M. Brisbane (ed.) (Oxford: Oxbow Books, forthcoming).

3. This long-running Institute project focuses on the Roman fort of Noviodunum on the south bank of the River Danube in Romania and its hinterland, see K. Lockyear, "At the edge of empires: the Noviodunum project, Romania" *Archaeology International* (2002/2003) 21–4; K. Lockyear, T. Sly and A. Popescu, "The Noviodunum Archaeological Project 2000–2004: results and conclusions from

the pilot seasons" *Peuce*, New Series **3–4** (2006–2007) 121–58.

4. I am indebted to Dr. K. Lockyear, Co-Director of the NAP, for the information on which this section is based.

5. See for example C. Orton, *Sampling in Archaeology*, 67–111 (Cambridge: Cambridge University Press, 2000).

6. This impinges on a long-running debate about the value of the concept of archaeological sites (see p. 81 in Orton 2000 n. 5 above, and references there). Those who do not like the terms can substitute "high density artefact areas" and "low density artefact areas", but for simplicity the terms "on site" and "off site'"will be used here.

7. A minor complication is that there may sometimes be fewer than 25 spots in a block, if natural circumstances make part of it inaccessible. Statistically, this is not an insuperable problem, but it needs to be kept in mind.

8. S.-D. Poisson, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*, (Paris: Bachelier, 1837), referred to by E. Parzen, *Modern Probability Theory and its Applications*, 255 (Tokyo: John Wiley & Sons, 1960).

9. L. von Bortkiewicz, *Das Gesetz der Kleinen Zahlen*. (Leipzig: Teubner, 1898).

10. R. D. Clarke, "An application of the Poisson distribution", *Journal of the Institute of Actuaries* **72,** 481, 1946; see also W. Feller, *An Introduction to Probability Theory and its Applications* Volume 1, 150 (New York: John Wiley & Sons, 1957).

11. D. O'Sullivan and D.J. Unwin, *Geographic Information Analysis*, 98–9 (Hoboken: John Wiley & Sons, 2003).

12. B.S. Everitt and D.J. Hand, *Finite Mixture Distributions*, 97–102 (London: Chapman and Hall, 1981). An older but more approximate method, the Method of Moments (MOM), is used to provide a starting point for the mle process.

13. This has been used to model the distribution of Yellow Cards in football matches, see J. Goddard, "Is the ref blind? Crime and Punishment in English Premier League football", *Significance* **4(2)**, 63–6, 2007.

14. This is partly due to the small number of spots per block (25). Larger numbers might require a mixture of three or more distributions.

15. These two models are often, but not always, the same. Sometimes a barely adequate fit can be dramatically improved by a more complex model. There is a trade-off between goodness-of-fit and complexity.